Computers and Mathematics with Applications 162 (2024) 52–60

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Adaptive least-squares finite element methods: Guaranteed upper bounds and convergence in L_2 norm of the dual variable ☆

JaEun Ku

Department of Mathematics, Oklahoma State University, 401 Mathematical Sciences, Stillwater, OK 74078, USA

ARTICLE INFO

Keywords:
Adaptive procedure
Least squares
Finite element methods

ABSTRACT

We consider adaptive least-squares finite element methods. First, we develop a guaranteed upper bound for the dual error in the L_2 norm, and this can be used as a stopping criterion for the adaptive procedures. Secondly, based on the a posteriori error estimates for the dual variable, we develop an error indicator that identifies the local area to refine, and establish the convergence of the adaptive procedures based on the Dörfler's marking strategy. Our convergence analysis is valid for the entire range of the bulk parameter $0 < \Theta \leq 1$ and it shows the effect of bulk parameter and reduction factor of elements on the convergence rate. Confirming numerical experiments are provided.

1. Introduction

Self-adaptive finite element methods have gained enormous importance for the numerical solution of partial differential equations. Typically, a posteriori error estimates are used to identify the local regions to refine the current discretization and repeat the process until the desired accuracy is achieved. There has been tremendous progress in developing error estimators and establishing convergence of adaptive procedures for the standard and mixed Galerkin finite element methods, see [1,3,9,11,13,15,17,19,23] and references therein.

Least-squares finite element methods (LSFEMs) provide competitive alternatives for computations of approximate solutions, see [4] and references therein. One of the main advantages of the LSFEMs is that they have a built-in a posteriori error estimator in the natural energy-type norms. However, there is only a limited number of results concerning the mathematical theory of adaptive least-squares finite element methods (ALSFEMs). Recently, convergence results concerning ALSFEMs are obtained in [7,8,10,11,14]. In particular, optimal convergence rates are obtained in [7,8,10,11]. These are significant advances. However, they are valid with some restrictions. For example, the results in [7,8,10,11] are restricted to the lowest-order Raviart-Thomas or Nédélec elements, and the results in [14] require a local bounded assumption. More importantly, the existing convergence results do not show the effect of bulk parameters and reduction rate of elements (ratio of new and current elements) on the convergence. It is expected and observed that there is a sharper decrease in the error between two consecutive levels when a large bulk parameter and small element reduction rate are used. There is no theoretical analysis justifying this for the ALSFEMs. Another important issue concerning ALSFEMs is that there are no guaranteed upper bounds for the ALSFEMs that can be used as a stopping criterion in the adaptive procedures. This can limit the application of ALSFEMs in practice. There is a result proving the asymptotic exactness of the least-squares functional and the error in $H(\mathrm{div}) \times H^1$ norm for the dual and primary variables, see [12].

There are two main goals in this paper. First, we develop a guaranteed upper bound for the dual error in the L_2 norm. This can be used as a stopping criterion in the adaptive procedures. Our numerical experiments show that the upper bounds are accurate and overestimate the actual error by a factor of less than 2. Secondly, we establish the convergence of ALSFEMs. A weighted version of a posteriori error estimators developed in [16] is proposed as an error indicator. With the weighted residual as an error indicator, convergence is established using the reduction of the indicator in the adaptive procedures based on the Dörfler marking strategy. Our approach establishes convergence of ALSFEMs in the $L_2 \times H^1$ norm of dual and primary variables. The argument can be easily extended to the convergence in the natural energy norm, i.e. $H(\mathrm{div}) \times H^1$ norm. As observed in [8], the difference between these two norms is that the first one uses data oscillation while the latter uses data approximation. Our analysis is valid for the entire range of the bulk parameter $0 < \Theta \leq 1$ and it shows that a larger bulk parameter Θ and small reduction rate elements γ result in a sharper decrease in the error between two consecutive levels in the adaptive procedure.

☆ This research is supported in part by NSF Grant DMS-2208289.
E-mail address: jku@okstate.edu.

https://doi.org/10.1016/j.camwa.2024.03.002
Received 3 August 2023; Received in revised form 11 February 2024; Accepted 1 March 2024
Available online 14 March 2024

More precisely, we show that for any given $\epsilon > 0$, there exists N_0 such that

$$\|u - u_h^{N_0+k}\|_1^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h^{N_0+k}\|_0^2 \le \epsilon + C\delta^k, \tag{1.1}$$

for $k = 1, 2, ...$, where $0 < \delta = \sqrt{1 - \Theta(1-\gamma^2)} < 1$, and $(u_h^n, \boldsymbol{\sigma}_h^n)$ is the approximate solution at the n-th step. Thus, one can expect that a linear rate of convergence, modulo ϵ, from N_0-step. While we prove the existence of such N_0, we cannot determine the actual value since it depends on many quantities such as ϵ and shape regularity of the elements etc.

One can use the marking strategy for weighted residuals in this paper to complement the existing strategies for the convergence of ALSFEMs. It is easy and simple to implement the marking strategy for the weighted residual along with the existing ones, and the convergence of the adaptive procedure is guaranteed from the convergence result of the weighted residual. For example, ALSFEMs in [11] converge at an optimal rate with the lowest-order Raviart-Thomas elements, the results do not extend to higher-order elements. By incorporating the marking strategy in this paper, convergence is guaranteed with higher-order elements.

This paper is organized as follows. Section 2 introduces the second-order elliptic problems and the least-squares formulation of the problem. The finite element approximations and some preliminary results are presented in Section 3. In Section 4, we establish an upper bound for the dual error. In Section 5, we show that the sequence of the approximate solutions is a Cauchy sequence in a Hilbert space and Section 6 presents a posteriori error indicators and mesh refinement strategy to guarantee convergence to the true solutions. Numerical examples are presented in Section 7.

2. Problem formulation

Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a bounded polygonal domain with boundary $\partial\Omega$. Let $H^s(D)$ denote the Sobolev space of order s defined on D, and the norm and semi-norm in $H^s(D)$ are denoted by $\|\cdot\|_{s,D}$ and $|\cdot|_{s,D}$ respectively. When $D = \Omega$, we will use $\|\cdot\|_s$, and $H_0^1(D)$ denote the functions in $H^1(D)$ with zero trace on ∂D. We shall also use the space

$$H(\mathrm{div}) = \{\boldsymbol{\tau} \in (L^2(\Omega))^d : \nabla\cdot\boldsymbol{\tau} \in L^2(\Omega)\},$$

with the norm $\|\boldsymbol{\tau}\|_{H(\mathrm{div})}^2 = (\nabla\cdot\boldsymbol{\tau}, \nabla\cdot\boldsymbol{\tau}) + (\boldsymbol{\tau}, \boldsymbol{\tau})$.

2.1. Second-order elliptic problems

We consider the following model elliptic partial differential equation:

$$-\nabla\cdot A\nabla u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{2.1}$$

where $f \in L_2(\Omega)$ and $A = (a_{ij}(x))_{i,j=1}^d$. We assume that a_{ij} is Lipschitz continuous and the matrix A is symmetric and uniformly positive definite, i.e. there exist constants α_0 and α_1 satisfying

$$\alpha_0\|\mathbf{x}\|^2 \le \mathbf{x}^T A\mathbf{x} \le \alpha_1\|\mathbf{x}\|^2, \text{ for all } \mathbf{x} \in \mathbb{R}^d. \tag{2.2}$$

We assume that $u \in H^2$ with the following regularity estimate:

$$\|u\|_2 \le C\|f\|_0. \tag{2.3}$$

Here and hereafter, we use C to denote a generic positive constant, that is, a constant independent of the mesh and f, but that may depend on the domain Ω.

Remark 2.1. We assume $u \in H^2(\Omega)$ for a simple presentation. If $u \in H^{1+\alpha}(\Omega)$ with $0 < \alpha < 1$, then the weight of the error indicator defined in (3.13) becomes $h_T^{2\alpha}$ instead of h_T^2, where h_T is the diameter of T.

2.2. Least-squares variational problems

The second-order equation is transformed into a system of first order by introducing dual variable $\boldsymbol{\sigma} = -A\nabla u$. Then, (2.1) becomes

$$\nabla\cdot\boldsymbol{\sigma} = f \quad \text{and} \quad \boldsymbol{\sigma} + A\nabla u = 0. \tag{2.4}$$

Let

$$\mathbf{V} = H(\mathrm{div}) \text{ and } S = H_0^1(\Omega)$$

Then, the least squares method for the first-order system (2.4) is: Find $u \in S, \boldsymbol{\sigma} \in \mathbf{V}$ such that

$$b(\boldsymbol{\sigma}, u; \boldsymbol{\tau}, v) = F(\boldsymbol{\tau}, v), \text{ for all } (\boldsymbol{\tau}, v) \in \mathbf{V}\times S, \tag{2.5}$$

where

$$b(\boldsymbol{\sigma}, u; \boldsymbol{\tau}, v) = (\nabla\cdot\boldsymbol{\sigma}, \nabla\cdot\boldsymbol{\tau}) + (A^{-1}(\boldsymbol{\sigma} + A\nabla u), \boldsymbol{\tau} + A\nabla v)$$

and

$$F(\boldsymbol{\tau}, v) = (f, \nabla\cdot\boldsymbol{\tau}).$$

3. Finite element approximation

To approximate the solution of (2.5), let $\mathcal{T}_h$ be a partition of Ω into triangles or rectangles (or their higher-dimensional analogues). For simplicity of presentation, we consider only triangular elements. We assume that the triangulation $\mathcal{T}_h$ is regular [5]. Let h_T be the diameter of element $T \in \mathcal{T}_h$ and let $h(x) = h_T$, where $x \in T \in \mathcal{T}_h$, denote a mesh function from Ω to $[0, 1]$.

Let $P_k(T)$ be the space of polynomials of degree k on triangle T. For theoretical analysis, consider

$$Q_h = \{q \in L_2(\Omega) : q|_T \in P_k(T), \text{ for each } T \in \mathcal{T}_h\}. \tag{3.1}$$

Let $P_h : L_2(\Omega) \to Q_h$ be local L_2 projection satisfying

$$(v - P_h v, q_h) = 0, \quad \forall q_h \in Q_h. \tag{3.2}$$

The projection operator P_h satisfies the following approximation property:

$$\|v - P_h v\|_{0,T} \le Ch_T^s |v|_{s,T}. \tag{3.3}$$

For the dual approximation spaces, we use the Raviart-Thomas spaces [21] or Brezzi-Douglas-Marini spaces [6]. For simplicity, we present our results based on the Raviart-Thomas spaces. The key requirement for the dual approximation spaces is the commuting diagram property given in (3.4). Denote the local Raviart-Thomas space of order k:

$$RT_k(T) = P_k(T)^d + \mathbf{x}P_k(T)$$

with $\mathbf{x} = (x_1, ..., x_d)$. The standard $H(\mathrm{div})$ conforming Raviart-Thomas space of index k is defined by

$$\mathbf{V}_h = \{\boldsymbol{\tau} \in \mathbf{V} : \boldsymbol{\tau}|_T \in RT_k(T) \text{ for all } T \in \mathcal{T}_h\}.$$

Then, we require that there exists an interpolant $\Pi_h : \mathbf{V} \cap [L^p(\Omega)]^n \to \mathbf{V}_h$, for some fixed $p > 2$, satisfying the following commuting diagram property.

$$\nabla\cdot\Pi_h\boldsymbol{\tau} = P_h\nabla\cdot\boldsymbol{\tau}, \tag{3.4}$$

for any $\boldsymbol{\tau} \in \mathbf{V}$. With the above commuting diagram property with (3.2), we have

$$(\nabla\cdot(\boldsymbol{\tau} - \Pi_h\boldsymbol{\tau}), v_h) = (\nabla\cdot\boldsymbol{\tau} - P_h\nabla\cdot\boldsymbol{\tau}, v_h) = 0, \tag{3.5}$$

for all $v_h \in Q_h$. For approximation properties, we have

$$\|\boldsymbol{\tau}-\Pi_h\boldsymbol{\tau}\|_{0,T}\le Ch_T^s|\boldsymbol{\tau}|_{s,T},\ \text{for}\ \frac{1}{2}<s\le k+1. \tag{3.6}$$

For the approximation spaces for the primary variable, we use the standard continuous piecewise polynomial spaces S_h defined as

$$S_h=\{\mathrm{v}\in S:\mathrm{v}|_T\in P_{k+1}(T)\ \text{for all}\ T\in\mathcal{T}_h\}.$$

It has the following approximation property [22]:

$$\|u-u_I\|_{1,T}\le Ch_T^{l-1}|u|_{l,T},\ \text{for}\ 1\le l\le k+2. \tag{3.7}$$

Let $\mathbf{V}_h\times S_h\subset\mathbf{V}\times S$ be the finite element space to approximate the dual variable $\boldsymbol{\sigma}$ and primary function u.

3.1. Finite element approximation

We define an approximate solution $(\boldsymbol{\sigma}_h,u_h)\in\mathbf{V}_h\times S_h$ for $(\boldsymbol{\sigma},u)$ in (2.5) as

$$b(\boldsymbol{\sigma}_h,u_h;\boldsymbol{\tau}_h,v_h)=F(\boldsymbol{\tau}_h,v_h),\quad\forall(\boldsymbol{\tau}_h,v_h)\in\mathbf{V}_h\times S_h. \tag{3.8}$$

Then, we have the following orthogonality property by subtracting (3.8) from (2.5):

$$b(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{u}-\mathbf{u}_h;\boldsymbol{\tau}_h,\mathrm{v}_h)=0,\ \text{for all}(\boldsymbol{\tau}_h,\mathrm{v}_h)\in\mathbf{V}_h\times S_h. \tag{3.9}$$

Using the definition of the bilinear form $b(\cdot;\cdot)$, we have

$$\begin{aligned}b(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,u-u_h;\boldsymbol{\tau}_h,v_h)&=(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),\nabla\cdot\boldsymbol{\tau}_h)\\&\quad+(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\boldsymbol{\tau}_h+A\nabla v_h)\\&=0,\quad\forall(\boldsymbol{\tau}_h,v_h)\in\mathbf{V}_h\times S_h.\end{aligned} \tag{3.10}$$

Taking $\boldsymbol{\tau}_h=0$ and using $\boldsymbol{\sigma}+A\nabla u=0$ in the above, we have

$$(\boldsymbol{\sigma}_h+A\nabla u_h,\nabla v_h)=0,\ \text{for all}\ v_h\in S_h. \tag{3.11}$$

The inequality (3.11) plays an important role in the development of a posteriori error indicators in this paper. It is well-known that the approximate solution $(\boldsymbol{\sigma}_h,\mathbf{u}_h)$ satisfies the minimization property.

3.2. Error estimators

To present our error estimators and indicators, first define

$$\eta^2(T)=\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_{0,T}^2, \tag{3.12}$$

$$\zeta^2(T)=h_T^2\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_{0,T}^2, \tag{3.13}$$

$$\mathrm{osc}^2(f,T)=\|h_T(f-P_hf)\|_{0,T}^2. \tag{3.14}$$

For any subset $\mathcal{M}\subset\mathcal{T}_h$, we define

$$\eta^2(\mathcal{M})=\sum_{T\in\mathcal{M}}\eta^2(T),$$

and $\zeta^2(\mathcal{M})$ and $\mathrm{osc}^2(f,\mathcal{M})$ are defined similarly.

Remark 3.1. For the remainder of this paper, for any $g\in L^2(\Omega)$ and a triangulation $\mathcal{T}_h$, set

$$\|hg\|_0=\Big(\sum_{T\in\mathcal{T}_h}h_T^2\|g\|_{0,T}^2\Big)^{1/2},$$

where h is a mesh function with $h(x)=h_T$ for $x\in T\in\mathcal{T}_h$.

In [16], the following inequality is obtained:

Theorem 3.1. *Let $(\boldsymbol{\sigma}_h,u_h)$ be the approximate solution defined in (3.8), and let $f_h=\nabla\cdot\boldsymbol{\sigma}_h$. Then,*

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0^2+\|\nabla(u-u_h)\|_0^2\le C\Big(\sum_{T\in\mathcal{T}_h}\eta^2(T)+\|h(f-f_h)\|_0^2\Big). \tag{3.15}$$

The following lower bound (efficiency bound) is obtained in [20, Theorem 5.2].

Theorem 3.2. *Assume that there exists $s\in\mathbf{N}$ such that $(A\nabla u_h)_T$ belongs to $P_s(T)^n$, for all $T\in\mathcal{T}_h$. Then, the following local lower bound holds:*

$$h_T\|\boldsymbol{\sigma}_h+A\nabla u_h\|_{0,T}\le C(\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,T}+\|u-u_h\|_{0,T}).$$

4. Guaranteed upper bounds

We develop an upper bound for the dual error $\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0$ and $\|A^{1/2}\nabla(u-u_h)\|_0$. We use the following Poincaré inequality, see [2].

$$\|v-P_hv\|_{0,T}\le\frac{h_T}{\pi}\|\nabla v\|_{0,T}, \tag{4.1}$$

and Friedrichs inequality

$$\|v\|_0\le\mathrm{diam}(\Omega)\|\nabla v\|_0. \tag{4.2}$$

Theorem 4.1. *Let $(\boldsymbol{\sigma}_h,u_h)$ be the approximation of (3.8). Then,*

$$\begin{aligned}\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\le\Big(&\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0^2\\&+\frac{2}{\alpha_0^2\pi^2}\|h(f-P_hf)\|_0^2\\&+2(\frac{\mathrm{diam}(\Omega)}{\alpha_0})^2\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2\Big)^{1/2},\end{aligned} \tag{4.3}$$

and

$$\begin{aligned}\|A^{1/2}\nabla(u-u_h)\|_0\le\Big(&\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0\\&+\frac{1}{\alpha_0\pi}\|h(f-P_hf)\|_0\\&+(\frac{\mathrm{diam}(\Omega)}{\alpha_0})\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\Big).\end{aligned} \tag{4.4}$$

Remark 4.1. The last two terms $\|h(f-P_hf)\|_0$ and $\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0$ are higher order terms. Clearly, $\|h(f-P_hf)\|_0$ is higher order compared to $\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0$. In section 6, we show that $\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\le\|h(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0$.

Proof. Using the integration by parts, $\boldsymbol{\sigma}+A\nabla u=0$ and Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2=(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\\&=(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)+(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)\\&=(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h))\\&\quad+(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)-((\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\nabla(u-u_h))\\&=(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h))\\&\quad+(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)-(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\nabla(u-u_h))-\|A^{1/2}\nabla(u-u_h)\|_0^2\\&=(A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)),\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h))\\&\quad+2(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)-\|A^{1/2}\nabla(u-u_h)\|_0^2\\&=\|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0^2-\|A^{1/2}\nabla(u-u_h)\|_0^2\\&\quad+2(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h).\end{aligned} \tag{4.5}$$

We have

$$\begin{aligned}(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)&=(\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma}),u-u_h)+(\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h),u-u_h)\\&=I_1+I_2.\end{aligned} \tag{4.6}$$

Now, using the orthogonal property (3.2) and approximation property (3.3), we have

$$
\begin{aligned}
I_1 &= (\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma}_h), u-u_h-P_h(u-u_h))\\
&\le \sum_T \|\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma}_h)\|_{0,T}\|u-u_h-P_h(u-u_h)\|_{0,T}\\
&\le \sum_T \|\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma}_h)\|_{0,T}\frac{h_T}{\pi}\|\nabla(u-u_h)\|_{0,T}\\
&\le \frac{1}{\pi}\|h(f-P_hf)\|\|\nabla(u-u_h)\|_0\\
&\le \frac{1}{\pi}\|h(f-P_hf)\|\frac{1}{\alpha_0}\|A^{1/2}\nabla(u-u_h)\|_0\\
&\le \frac{1}{\alpha_0^2\pi^2}\|h(f-P_hf)\|_0^2+\frac{1}{4}\|A^{1/2}\nabla(u-u_h)\|_0^2.
\end{aligned}
$$

For I_2, using Friedrichs inequality, we have

$$
\begin{aligned}
I_2 &\le \|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\|u-u_h\|_0\\
&\le \mathrm{diam}(\Omega)\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\|\nabla(u-u_h)\|_0\\
&\le \frac{\mathrm{diam}(\Omega)}{\alpha_0}\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\|A^{1/2}\nabla(u-u_h)\|_0\\
&\le (\frac{\mathrm{diam}(\Omega)}{\alpha_0})^2\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2+\frac{1}{4}\|\|A^{1/2}\nabla(u-u_h)\|_0^2.
\end{aligned}
$$

Now, plugging the inequalities for I_1 and I_2 into (4.6) and then (4.5), we have

$$
\begin{aligned}
&\|A^{-1/2}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2\\
&\le \|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0^2+\frac{2}{\alpha_0^2\pi^2}\|h(f-P_hf)\|_0^2\\
&\quad +2(\frac{\mathrm{diam}(\Omega)}{\alpha_0})^2\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2.
\end{aligned}
$$

Taking square root on both sides, we obtain (4.3).

Similarly, we have

$$
\begin{aligned}
\|A^{1/2}\nabla(u-u_h)\|_0^2 &= (A\nabla(u-u_h)\,,\,\nabla(u-u_h))\\
&= (A^{-1}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h+A\nabla(u-u_h)), A\nabla(u-u_h))\\
&\quad +(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h), u-u_h)\\
&\le \|A^{-1/2}(\boldsymbol{\sigma}_h+A\nabla u_h)\|_0\|A^{1/2}\nabla(u-u_h)\|_0\\
&\quad +(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h), u-u_h).
\end{aligned}
$$

Now, using the argument for I_1 and I_2 we have

$$
\begin{aligned}
&(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h), u-u_h)\\
&\le \Big(\frac{1}{\alpha_0\pi}\|h(f-P_hf)\|_0+\frac{\mathrm{diam}(\Omega)}{\alpha_0}\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0\Big)\\
&\quad\times\|A^{1/2}\nabla(u-u_h)\|_0.
\end{aligned}
$$

Finally, using the above two inequalities and canceling $\|A^{1/2}\nabla(u-u_h)\|_0$, we obtain (4.4). This completes the proof. □

5. Convergence of the approximate solutions $\{(\sigma_h^n, u_h^n)\}_{n=1}^\infty$

Let

$$\mathbf{X} = \mathbf{V}\times S.$$

For notational convenience, we use $\mathcal{U}=(\boldsymbol{\tau}, v)\in\mathbf{X}$ and define the energy-type norm for $\mathcal{U}=(\boldsymbol{\tau}, v)$ as follows:

$$|||\mathcal{U}|||^2 = b(\boldsymbol{\tau}, v; \boldsymbol{\tau}, v). \tag{5.1}$$

It is well-known that there exists $C>0$ independent of meshsize satisfying

$$\frac{1}{C}(\|\boldsymbol{\tau}\|_{H(\mathrm{div})}^2+\|v\|_1^2)\le |||\mathcal{U}|||^2\le C(\|\boldsymbol{\tau}\|_{H(\mathrm{div})}^2+\|v\|_1^2), \tag{5.2}$$

Let $\mathbf{X}_h^n=\mathbf{V}_h^n\times S_h^n, n=1,2,3,\dots$ be the nested sequence of approximation spaces on $\mathcal{T}_h^n$, and $\mathcal{U}_h^n=(\boldsymbol{\sigma}_h^n, u_h^n)\in\mathbf{X}_h^n$ be the approximate solution defined by

$$B(\mathcal{U}_h^n, \mathcal{V}_h^n) = F(\mathcal{V}_h),\ \text{ for all }\ \mathcal{V}_h^n=(\boldsymbol{\tau}_h^n, v_h^n)\in\mathbf{X}_h^n, \tag{5.3}$$

where

$$B(\mathcal{U}_h^n, \mathcal{V}_h^n) = b(\boldsymbol{\sigma}_h^n, u_h^n; \boldsymbol{\tau}_h^n, v_h^n)\quad\text{and}\quad F(\mathcal{V}_h^n)=F(\boldsymbol{\tau}_h^n, v_h^n).$$

Then, the orthogonality property (3.9) can be written as

$$B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{V}_h^n)=0,\ \text{for all } \mathcal{V}_h^n\in\mathbf{X}_h^n, \tag{5.4}$$

where $\mathcal{U}=(\boldsymbol{\sigma}, \mathbf{u})$ in (2.5).

Lemma 5.1. *For $n=1,2,3,\dots$, let $\mathcal{U}_h^n\in\mathbf{X}_h$ be the solution defined in (5.3), and $\mathcal{U}\in\mathbf{X}$ be the solution of (2.5). Then, for $m>n$*

$$|||\mathcal{U}_h^n-\mathcal{U}_h^m|||^2 = \sum_{i=n}^{m-1}|||\mathcal{U}_h^i-\mathcal{U}_h^{i+1}|||^2. \tag{5.5}$$

Proof. For $m>n$, using the orthogonality (5.4), we have

$$
\begin{aligned}
|||\mathcal{U}_h^m-\mathcal{U}_h^n|||^2 &= B(\mathcal{U}_h^m-\mathcal{U}_h^n, \mathcal{U}_h^m-\mathcal{U}_h^n)\\
&= B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}_h^m-\mathcal{U}_h^n)\\
&= B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}_h^m-\mathcal{U})+B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}-\mathcal{U}_h^n)\\
&= -B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}-\mathcal{U}_h^m)+B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}-\mathcal{U}_h^n)\\
&= -B(\mathcal{U}-\mathcal{U}_h^m, \mathcal{U}-\mathcal{U}_h^m)+B(\mathcal{U}-\mathcal{U}_h^n, \mathcal{U}-\mathcal{U}_h^n)\\
&= -|||\mathcal{U}-\mathcal{U}_h^m|||^2+|||\mathcal{U}-\mathcal{U}_h^n|||^2.
\end{aligned}
$$

Rearranging the terms, we obtain

$$|||\mathcal{U}-\mathcal{U}_h^m|||^2 = |||\mathcal{U}-\mathcal{U}_h^n|||^2-|||\mathcal{U}_h^m-\mathcal{U}_m^n|||^2. \tag{5.6}$$

Now, using (5.6) repeatedly, we have

$$
\begin{aligned}
|||\mathcal{U}-\mathcal{U}_h^{n+1}|||^2 &= |||\mathcal{U}-\mathcal{U}_h^n|||^2-|||\mathcal{U}^n-\mathcal{U}_h^{n+1}|||^2,\\
|||\mathcal{U}-\mathcal{U}_h^{n+2}|||^2 &= |||\mathcal{U}-\mathcal{U}_h^{n+1}|||^2-|||\mathcal{U}^{n+1}-\mathcal{U}_h^{n+2}|||^2,\\
&\vdots\\
|||\mathcal{U}-\mathcal{U}_h^{m-1}|||^2 &= |||\mathcal{U}-\mathcal{U}_h^{m-2}|||^2-|||\mathcal{U}^{m-2}-\mathcal{U}_h^{m-1}|||^2,\\
|||\mathcal{U}-\mathcal{U}_h^{m}|||^2 &= |||\mathcal{U}-\mathcal{U}_h^{m-1}|||^2-|||\mathcal{U}^{m-1}-\mathcal{U}_h^{m}|||^2.
\end{aligned}
$$

Adding the above equations and canceling the same terms, we obtain

$$|||\mathcal{U}-\mathcal{U}_h^m|||^2 = |||\mathcal{U}-\mathcal{U}_h^n|||^2-\sum_{i=n}^{m-1}|||\mathcal{U}_h^i-\mathcal{U}_h^{i+1}|||^2.$$

Now, using (5.6) with the above equality, we obtain (5.5). This completes the proof. □

Theorem 5.2. *For $n=1,2,3,\dots$, let $\mathcal{U}_h^n=(\boldsymbol{\sigma}_h^n, u_h^n)\in\mathbf{X}_h^n$ be the solution defined in (5.3). Then, $\{\mathcal{U}_h^n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbf{X}$, and*

$$\lim_{n\to\infty}\mathcal{U}_h^n=\mathcal{U}^\infty,\ \text{for some } \mathcal{U}^\infty=(\boldsymbol{\sigma}^\infty, u^\infty)\in\mathbf{X}=\mathbf{V}\times S,$$

i.e.

$$\boldsymbol{\sigma}_h^n\to\boldsymbol{\sigma}^\infty \text{ in } H(\mathrm{div}),\ \text{and } u_h^n\to u^\infty \text{ in } H_0^1(\Omega).$$

Proof. Define a sequence a_i and S_i as

$$a_i = |||\mathcal{U}_h^i-\mathcal{U}_h^{i+1}|||^2,\quad S_i=\sum_{j=1}^i a_i,\ \text{for } i=1,2,3,\dots$$

Note that using (5.5),

$$S_i = \sum_{j=1}^{i} a_j = \sum_{j=1}^{i} |||\mathcal{U}_h^j - \mathcal{U}_h^{j+1}|||^2 = |||\mathcal{U}_h^1 - \mathcal{U}_h^{i+1}|||^2.$$

Now, using (5.6) with $m = i+1, n = 1$ and using the above equality, we have

$$|||\mathcal{U} - \mathcal{U}_h^{i+1}|||^2 = |||\mathcal{U} - \mathcal{U}_h^1|||^2 - |||\mathcal{U}_h^1 - \mathcal{U}_h^{i+1}|||^2 = |||\mathcal{U} - \mathcal{U}_h^1|||^2 - S_i.$$

Thus, $\{S_i\}_{i=1}^{\infty}$ is nonnegative, increasing, and bounded above with an upper bound $|||\mathcal{U} - \mathcal{U}_h^1|||^2$. Hence, $\{S_i\}_{i=1}^{\infty}$ converges, and this implies that $\{S_i\}_{i=1}^{\infty}$ is a Cauchy sequence.

Now, to show that $\{\mathcal{U}_h^n\}_{n=1}^{\infty}$ is a Cauchy sequence, let $\delta > 0$ be given. Then, there exists $N > 0$ such that $|S_l - S_k| < \delta^2$ for $k, l, > N$ since $\{S_i\}_{i=1}^{\infty}$ is a Cauchy sequence. Without loss of generality, $l > k$. Now, $|S_l - S_k| = |||\mathcal{U}_{k+1} - \mathcal{U}_{l+1}|||^2$. Thus,

$$|||\mathcal{U}_{k+1} - \mathcal{U}_{l+1}||| < \delta, \text{ for } k, l > N.$$

This implies that $\{\mathcal{U}_h^n\}_{n=1}^{\infty} \subset \mathbf{X}$ is a Cauchy sequence. Thus, the Cauchy sequence $\{\mathcal{U}_h^n\}_{n=1}^{\infty}$ converges since $\mathbf{X} = H(\mathrm{div}) \times H_0^1(\Omega)$ is a Banach space, see [5]. We denote the limit as $\mathcal{U}^{\infty}$. This completes the proof. □

6. A posteriori estimates

In this section, we establish an upper bound for $\|A^{-1/2}(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0^2$, and this will be used to develop an error indicator to mark the local region to refine in the mesh refinement strategy.

6.1. Error indicators

We use the a posteriori error indicator ζ defined in (3.13) for a marking of current mesh, and show that the indicator is an upper bound for the error of the LS approximate solutions, and converges to 0 in the adaptive procedures.

Theorem 6.1. *Let $(\boldsymbol{\sigma}_h, u_h)$ be the approximate solution for (2.5) and Π_h be the Fortin interpolant in $\mathbf{V}_h$. Then, for $r = 1, 2$*

$$\|A^{-1/2}(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0^2 + \|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^r \le C\,\|u\|_2 \cdot \|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0, \tag{6.1}$$

Remark 6.1. From the above inequality and definitions (3.12) and (3.13), we have

$$\eta^2(\mathcal{T}_h) \le C\,\|u\|_2\,\zeta(\mathcal{T}_h). \tag{6.2}$$

Proof. Using the commuting diagram property (3.4), orthogonality (3.10), $\boldsymbol{\sigma} + A\nabla u = 0$ and A being uniformly positive definite, we have

$$\begin{aligned}
&\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^2 = (\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h))\\
&= (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h))\\
&= -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\\
&= -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma} + \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\\
&= -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma})\\
&\quad -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\\
&= -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma})\\
&\quad -(A^{-1}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h)), \boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h))\\
&\quad +(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h + A\nabla(u - u_h), \nabla(u - u_h))\\
&= (\boldsymbol{\sigma}_h + A\nabla u_h, \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}) - \|A^{-1/2}(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0^2\\
&\quad -(\boldsymbol{\sigma}_h + A\nabla u_h, \nabla(u - u_h)).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^2 + \|A^{-1/2}(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0^2\\
&\quad = (\boldsymbol{\sigma}_h + A\nabla u_h, \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}) - (\boldsymbol{\sigma}_h + A\nabla u_h, \nabla(u - u_h)).
\end{aligned}$$

Now, using (3.11), the approximation properties (3.6) and (3.7), we have

$$\begin{aligned}
&\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^2 + \|\boldsymbol{\sigma}_h + \nabla u_h\|_0^2\\
&= (\boldsymbol{\sigma}_h + A\nabla u_h, \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}) - (\boldsymbol{\sigma}_h + A\nabla u_h, \nabla(u - u_h))\\
&= (\boldsymbol{\sigma}_h + A\nabla u_h, \Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}) - (\boldsymbol{\sigma}_h + A\nabla u_h, \nabla(u - u_I))\\
&\le \sum_{T\in\mathcal{T}_h} \|\boldsymbol{\sigma}_h + A\nabla u_h\|_{0,T}\|\boldsymbol{\sigma} - \Pi_h\boldsymbol{\sigma}\|_{0,T}\\
&\quad + \sum_{T\in\mathcal{T}_h} \|\boldsymbol{\sigma}_h + A\nabla u_h\|_{0,T}\|u - u_I\|_{1,T}\\
&\le \sum_{T\in\mathcal{T}_h} h_T\|\boldsymbol{\sigma}_h + A\nabla u_h\|_{0,T}(\|\boldsymbol{\sigma}\|_{1,T} + \|u\|_{2,T})\\
&\le C(\|\boldsymbol{\sigma}\|_1 + \|u\|_2)\|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0\\
&\le C\|u\|_2\|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0.
\end{aligned}$$

This proves for $r = 2$.

Now, in order to reduce the power of $\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^2$ to $\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0$, consider the following auxiliary problem:

$$-\nabla\cdot A\nabla w = \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h), \qquad \boldsymbol{\eta} = -A\nabla w$$

Then, $\nabla\cdot\boldsymbol{\eta} = \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)$ with $\|\boldsymbol{\eta}\|_1 \le C\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0$. Using the above, we have

$$\begin{aligned}
&\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0^2\\
&= (\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h))\\
&= (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h))\\
&= (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot\boldsymbol{\eta})\\
&= (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h),\, \nabla\cdot(\boldsymbol{\eta} - \Pi_h\boldsymbol{\eta})) - (A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), \Pi_h\boldsymbol{\eta})\\
&= -(A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), \Pi_h\boldsymbol{\eta}) = -(A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), \Pi_h\boldsymbol{\eta} - \boldsymbol{\eta} + \boldsymbol{\eta})\\
&= -(A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), \Pi_h\boldsymbol{\eta} - \boldsymbol{\eta}) - (A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), -\nabla w)\\
&= -(A^{-1}(\boldsymbol{\sigma}_h + A\nabla u_h), \Pi_h\boldsymbol{\eta} - \boldsymbol{\eta}) + ((\boldsymbol{\sigma}_h + A\nabla u_h), \nabla w - \nabla w_I)\\
&\le C\|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0\cdot(\|\boldsymbol{\eta}\|_1 + \|w\|_2)\\
&\le C\|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0\|\nabla\cdot(\Pi_h\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_0
\end{aligned}$$

This completes the proof for (6.1). □

We present one of the main results in this paper.

Theorem 6.2. *Let $(\boldsymbol{\sigma}_h, u_h)$ be the approximate solution for (2.5) and Π_h be the Fortin interpolant in $\mathbf{V}_h$. Then,*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0^2 + \|\nabla(u - u_h)\|_0^2 \le C\,\|u\|_2\,\zeta(\mathcal{T}_h) + C\mathrm{osc}^2(f, \mathcal{T}_h). \tag{6.3}$$

Remark 6.2. The following efficiency bound is presented in Theorem 3.2, see [20, Theorem 5.2].

$$\zeta(T) \le \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,T} + \|u - u_h\|_{0,T}.$$

Proof. Using (3.15) and (6.2), we have

$$\begin{aligned}
\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0^2 + \|\nabla(u - u_h)\|_0^2 &\le C\eta^2(\mathcal{T}_h) + C\|h(f - f_h)\|_0^2\\
&\le C\,\|u\|_2\,\zeta(\mathcal{T}_h) + C\|h(f - f_h)\|_0^2.
\end{aligned} \tag{6.4}$$

Using $f_h = \nabla \cdot \boldsymbol{\sigma}_h$, the triangle inequality (3.14) and Theorem 6.1, we have

$$\begin{aligned}\|h(f-f_h)\|_0^2 &= \|h\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2\\ &\le C\|h\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma})\|_0^2 + C\|h\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2\\ &\le C\|h\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma})\|_0^2 + C\|\nabla\cdot(\Pi_h\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0^2\\ &\le C\|h\nabla\cdot(\boldsymbol{\sigma}-\Pi_h\boldsymbol{\sigma})\|_0^2 + C\|h(\boldsymbol{\sigma}_h + A\nabla u_h)\|_0\\ &= C\Big(\mathrm{osc}(f,\mathcal{T}_h)\Big)^2 + C\,\|u\|_2\,\zeta(\mathcal{T}_h).\end{aligned}$$

Now, plugging the above inequality into (6.4), we obtain (6.3). This completes the proof. □

6.2. Bulk parameter and refinement strategy

Let $\mathcal{T}_h^n$ be the current triangulation. For any subset $\mathcal{D} \subset \mathcal{T}_h^n$ we define

$$\zeta^2(\mathcal{D}) = \sum_{T\in\mathcal{D}} \zeta^2(T).$$

Based on Theorem 6.2, we first present a mesh refinement strategy for the weighted residual ζ and we provide a convergence analysis based on the strategy.

Marking strategy for weighted residual
Given a parameter $0 < \Theta \le 1$, construct a subset $\mathcal{M}_h^n$ of $\mathcal{T}_h^n$ such that

$$\Theta\,\zeta^2(\mathcal{T}_h^n) \le \zeta^2(\mathcal{M}_h^n). \tag{6.5}$$

Using $\zeta^2(\mathcal{T}_h^n) = \zeta^2(\mathcal{M}_h^n) + \zeta^2((\mathcal{M}_h^n)^C)$, then isolating $\zeta^2((\mathcal{M}_h^n)^C)$, we have

$$\Theta\,\zeta^2((\mathcal{M}_h^n)^C) \le (1-\Theta)\zeta^2(\mathcal{M}_h^n).$$

Now, adding $(1-\Theta)\zeta^2((\mathcal{M}_h^n)^C)$ on both sides of the above, we obtain

$$\zeta^2((\mathcal{M}_h^n)^C) \le (1-\Theta)\,\zeta^2(\mathcal{T}_h^n). \tag{6.6}$$

Concerning the relationship between two consecutive levels of mesh refinement, we assume the following: Let $\mathcal{T}_h^n$ and $\mathcal{T}_h^{n+1}$ be two consecutive triangulations in $\{\mathcal{T}_h^n\}_{n\ge 0}$ so that $\mathcal{T}_h^{n+1}$ is a refinement of $\mathcal{T}_h^n$. We assume that there exists a positive constant $0 < \gamma < 1$, the reduction rate of elements, satisfying

$$h_{T'} \le \gamma h_T \text{ if } T \in \mathcal{T}_h^n,\ T' \in \mathcal{T}_h^{n+1} \text{ and } T' \subset T. \tag{6.7}$$

Let h_n be the mesh function on $\mathcal{T}_h^n$ defined by

$$h_n(x) = h_T, \text{ where } x \in T \in \mathcal{T}_h^n. \tag{6.8}$$

Then, (6.7) becomes

$$\frac{h_{n+1}}{h_n} \le \gamma, \text{ for } x \in T \in \mathcal{M}_h^n. \tag{6.9}$$

6.3. Convergence

Now, we are ready to show the key ingredient for the convergence of adaptive procedures. The result states that ζ converges linearly modulo $\epsilon > 0$, and the results are valid for any bulk parameter $0 < \Theta \le 1$ and any order of approximation spaces.

Theorem 6.3. *Let $(\boldsymbol{\sigma}_h^n, u_h^n)$ be the approximate solutions of (2.5) on triangulations $\mathcal{T}_h^n, n = 0,1,2,3,\ldots$. Then, with the refinement strategy (6.5), we have*

$$\zeta(\mathcal{T}_h^n) = \Big(\sum_{T\in\mathcal{T}_h^n} h_T^2 \|\boldsymbol{\sigma}_h^n + A\nabla u_h^n\|_{0,T}^2\Big)^{1/2} \to 0, \text{ as } n\to\infty. \tag{6.10}$$

Moreover, for any $\epsilon > 0$, there exists $N_0 > 0$ such that

$$\zeta(\mathcal{T}_h^{N_0+k}) \le \epsilon + \delta^k \cdot \zeta(\mathcal{T}_h^{N_0}) \text{ for } k = 1,2,3,\ldots. \tag{6.11}$$

where

$$\delta = \sqrt{(1-\Theta(1-\gamma^2)}\ < 1. \tag{6.12}$$

Proof. Using Theorem 5.2, we have

$$\boldsymbol{\sigma}_h^n \to \boldsymbol{\sigma}^\infty \text{ and } u_h^n \to u^\infty.$$

Also, for each $x \in \Omega$, $\{h(x)\}_{n=0}^\infty$, with $0 \le h_n(x) \le 1$, is decreasing and bounded below by 0. Thus, $h_n(x)$ converges, and we denote the limit as $h_\infty(x)$. Then, we have

$$\boldsymbol{\sigma}_h^n + A\nabla u_h^n \to \boldsymbol{\sigma}^\infty + A\nabla u^\infty \text{ in } L^2, \text{ and } h_n(x) \to h_\infty(x) \text{ pointwise}$$

and $\|h_n\|_\infty \le 1$. Thus,

$$\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_0 \to \|h_\infty(\boldsymbol{\sigma}^\infty + A\nabla u^\infty)\|_0 = \beta, \text{ for some } \beta \ge 0. \tag{6.13}$$

Using the triangle inequality and $0 \le h_n \le 1$ for all n, we have

$$\begin{aligned}&\|h_{n+1}(\boldsymbol{\sigma}_h^{n+1} + A\nabla u_h^{n+1})\|_0\\ &\le \|h_{n+1}(\boldsymbol{\sigma}_h^{n+1} - \boldsymbol{\sigma}_h^n + A\nabla(u_h^{n+1} - u_h^n))\|_0 + \|h_{n+1}(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_0\\ &= E_n + \|h_{n+1}(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_0,\end{aligned} \tag{6.14}$$

where $E_n = \|h_{n+1}^\alpha(\boldsymbol{\sigma}_h^{n+1} - \boldsymbol{\sigma}_h^n + A\nabla(u_h^{n+1} - u_h^n))\|_0$. Now, using the refinement strategy satisfying $\frac{h_{n+1}}{h_n} \le \gamma < 1$ for the triangles marked for refinement and (6.6), i.e. $\|h_n(\boldsymbol{\sigma}_h^n + \nabla u_h^n)\|_{0,\mathcal{M}^C} \le (1-\Theta)\|h_n(\boldsymbol{\sigma}_h^n + \nabla u_h^n)\|_{0,\Omega}$, we have

$$\begin{aligned}&\|h_{n+1}(\boldsymbol{\sigma}_h^n + \mathcal{A}\nabla u_h^n)\|_0^2\\ &= \|h_{n+1}(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\mathcal{M}^C}^2 + \|\frac{h_{n+1}}{h_n} h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\mathcal{M}}^2\\ &\le \|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\mathcal{M}^C}^2 + \gamma^2\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\mathcal{M}}^2\\ &= (1-\gamma)\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\mathcal{M}^C}^2 + \gamma^2\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2\\ &\le (1-\gamma^2)(1-\Theta)\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2 + \gamma^2\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2\\ &= \Big((1-\gamma^2)(1-\Theta) + \gamma^2\Big)\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2\\ &= \Big((1-\Theta(1-\gamma^2)\Big)\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2\\ &= \delta^2\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_{0,\Omega}^2,\end{aligned}$$

where $0 \le \delta = \sqrt{1-\Theta(1-\gamma^2)} < 1$. Plugging the above inequality into (6.14), we have

$$\|h_{n+1}(\boldsymbol{\sigma}_h^{n+1} + A\nabla u_h^{n+1})\|_0 \le E_n + \delta\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_0. \tag{6.15}$$

To show that the limit converges to 0, using $0 \le h_n \le 1$ and convergence of $\{\boldsymbol{\sigma}_h^n\}, \{u_h^n\}$ we have

$$0 \le E_n \le \|\boldsymbol{\sigma}_h^{n+1} - \boldsymbol{\sigma}_h^n\|_0 + \|A(u_h^{n+1} - u_h^n)\|_0 \to 0 \text{ as } n\to\infty. \tag{6.16}$$

Taking $n \to \infty$ in (6.15) with using (6.13) and (6.16), we have

$$0 \le \beta \le \delta\cdot\beta, \text{ with } \delta = \sqrt{1-\Theta(1-\gamma^2)} < 1.$$

Thus, $\beta = 0$, i.e.

$$\|h_n(\boldsymbol{\sigma}_h^n + A\nabla u_h^n)\|_0 \to 0, \text{ as } n\to\infty.$$

This completes (6.10).

Now, using (6.16), i.e. $\lim_{n\to\infty} E_n = 0$, given $\epsilon > 0$, there exists N_0 such that

$$E_n < (1-\delta)\cdot\epsilon, \text{ for all } n > N_0. \tag{6.17}$$

Set

$$\zeta_N = \zeta(\mathcal{T}_h^N).$$

Now using (6.11) repeatedly and (6.17), we have

$$\begin{aligned}
\zeta_{N_0+k} &\le E_{N_0+k} + \delta\zeta_{N_0+k-1} \le E_{N_0+k} + \delta(E_{N_0+k-1} + \delta\zeta_{N_0+k-2}) \\
&\le E_{N_0+k} + \delta E_{N_0+k-1} + \delta^2\zeta_{N_0+k-2} \\
&\le \cdots \\
&\le E_{N_0+k} + \delta E_{N_0+k-1} + \cdots + \delta^{k-1}E_{N_0+1} + \delta^k\cdot\zeta_{N_0} \\
&\le (1-\delta)\epsilon(1+\delta+\delta^2+\cdots+\delta^{k-1}) + \delta^k\cdot\zeta_{N_0} \\
&\le (1-\delta)\epsilon\frac{1}{1-\delta} + \delta^k\cdot\zeta_{N_0} \\
&= \epsilon + \delta^k\cdot\zeta_{N_0}.
\end{aligned}$$

This proves (6.11). This completes the proof. □

Remark 6.3. With a larger bulk parameter Θ and a smaller refinement parameter γ, $\delta = \sqrt{1-\Theta(1-\gamma^2)}$ becomes smaller and indicates a sharper decrease in the error. Note that a larger bulk parameter means marking more triangles for refinement and smaller γ implies more refined meshes.

Note that $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0$ is bounded by ζ and the data oscillation $\mathrm{osc}(f,\cdot)$. Thus, we use the following marking strategy to ensure a decrease in the data oscillation.

Marking strategy for data oscillation
Given a parameter $0 < \Theta \le 1$ and subset $\mathcal{M}_h \subset \mathcal{T}_n$, enlarge $\mathcal{M}_n$ such that

$$\Theta\,\mathrm{osc}^2(f,\mathcal{T}_n) \le \mathrm{osc}^2(f,\mathcal{M}_n). \tag{6.18}$$

Remark 6.4. One can choose different bulk parameters for the weighted residual and the data oscillation. In this case, the convergence rate will be different. Also, in our numerical experiments, the marking strategy for data oscillation does not add significantly many elements to the existing $\mathcal{M}_n$ from the marking of the weighted residual.

With the bulk parameter Θ and reduction parameter γ, the following reduction of the data oscillation is well known, e.g. [18, Lemma 3.9]

$$\mathrm{osc}(f,\mathcal{T}_n) \le \delta^n \mathrm{osc}(f,\mathcal{T}_0) \text{ for } n = 0,1,2,\ldots \tag{6.19}$$

Theorem 6.4. *Let $\epsilon > 0$ be given. Then, there exists $N_0 > 0$ such that*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h^{N_0+k}\|_0^2 + \|\nabla(u - u_h^{N_0+k})\|_0^2 \le \epsilon + C\|f\|_0^2\delta^k, \tag{6.20}$$

where

$$\delta = \sqrt{1-\Theta(1-\gamma^2)}.$$

Proof. Using Theorem 3.1 and Theorem 6.3, and (6.19), we have

$$\begin{aligned}
&\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h^{N_0+k}\|_0^2 + \|\nabla(u - u_h^{N_0+k})\|_0^2 \\
&\le C\,\|u\|_2\,\zeta_{N_0+k} + C\Big(\mathrm{osc}(f,\mathcal{T}^{N_0+k})\Big)^2 \\
&\le C\,\|u\|_2\,\epsilon + C\,\|u\|_2\,\delta^k\zeta_{N_0} + C\delta^{2(N_0+k)}(\mathrm{osc}(f,\mathcal{T}^0))^2.
\end{aligned}$$

Now, using (2.2), we have

$$\|\boldsymbol{\tau}\|_0 \le C\|A^{-1/2}\boldsymbol{\tau}\|_0, \text{ for all } \boldsymbol{\tau},$$

and using (3.8), the arithmetic-geometric inequality, we have

$$\|A^{-1/2}(\boldsymbol{\sigma}_h^{N_0} + A\nabla u_h^{N_0})\|_0 \le \frac{1}{2}\|f\|_0.$$

Using $0 < h_n < 1$ and combining the above two inequalities, we have

$$\begin{aligned}
\zeta_{N_0} &= \|h_{N_0}(\boldsymbol{\sigma}_h^{N_0} + A\nabla u_h^{N_0})\|_0 \le \|(\boldsymbol{\sigma}_h^{N_0} + A\nabla u_h^{N_0})\|_0 \\
&\le C\|A^{-1/2}(\boldsymbol{\sigma}_h^{N_0} + A\nabla u_h^{N_0})\|_0 \\
&\le C\|f\|_0.
\end{aligned}$$

Clearly,

$$\mathrm{osc}(f,\mathcal{T}^0) \le \|f\|_0.$$

Thus, we have

$$\begin{aligned}
&\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h^{N_0+k}\|_0^2 + \|\nabla(u - u_h^{N_0+k})\|_0^2 \\
&\le C\,\|u\|_2\,\epsilon + C\Big(\|u\|_2\|f\|_0 + \delta^{2N_0+k}\|f\|_0^2\Big)\delta^k.
\end{aligned}$$

Now, scaling $C\|u\|_2\epsilon$ to ϵ, using $0 < \delta < 1$ and the regularity estimate (2.3), we obtain

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h^{N_0+k}\|_0^2 + \|\nabla(u - u_h^{N_0+k})\|_0^2 \le \epsilon + C\|f\|_0^2\,\delta^k.$$

This completes the proof. □

7. Numerical examples

In this section, we present numerical examples showing the convergence behavior of the ALSFEMs using the new error indicator $\zeta(\cdot)$ defined in (3.13) and data oscillation $\mathrm{osc}(f,\cdot)$ defined in (3.14), and

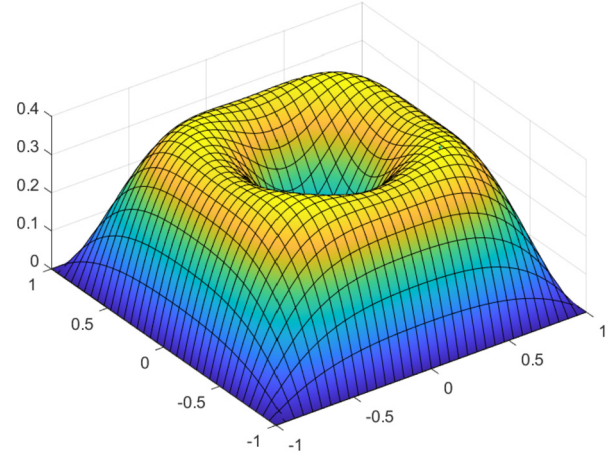

Fig. 7.1. A surface plot of the true solution u.

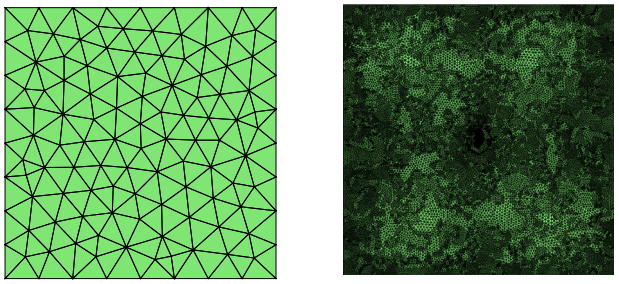

Fig. 7.2. Left: initial mesh with DoF = 2092; Right: generated by adaptive procedure with DoF = 82350.

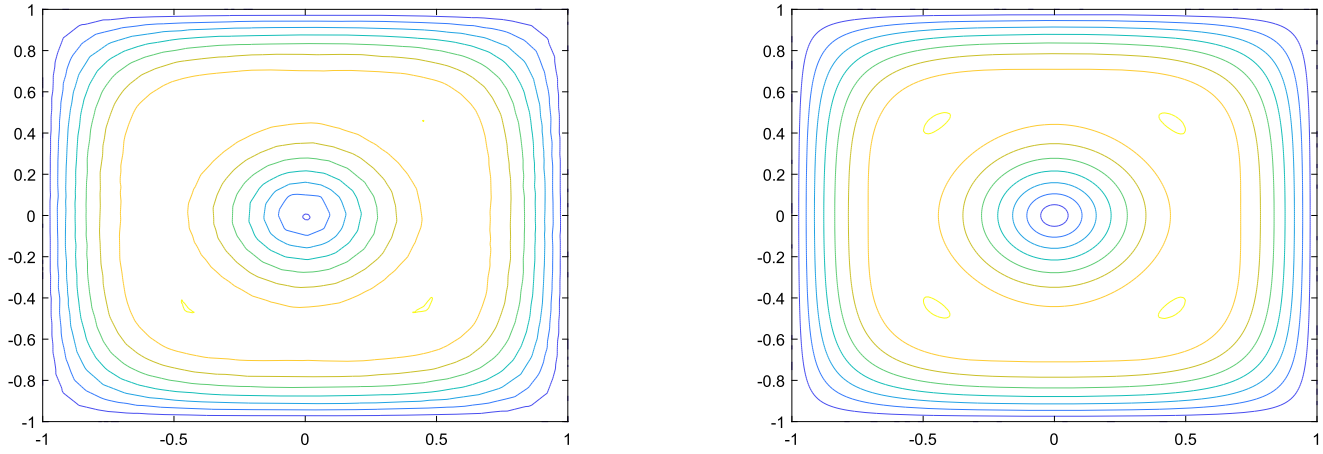

Fig. 7.3. Contour plots of u_h, Left: DoF = 2092; Right: DoF = 82350.

Table 7.1
Convergence and upper bounds with $\Theta = 0.8$ and $\delta = 0.6325$.

DoFs	$\|A^{-1/2}(\sigma - \sigma_h)\|_0$	rate	$\mathcal{UB}(\sigma)$	$\|A^{1/2}\nabla(u - u_h)\|_0$	rate	$\mathcal{UB}(u)$
2092	0.2171	-	0.3383	0.2338	-	0.4228
5201	0.1296	0.5970	0.1944	0.1494	0.6256	0.2255
13805	0.0832	0.6420	0.1241	0.0949	0.6352	0.1370
33693	0.0541	0.6502	0.0813	0.0605	0.6375	0.0906
82350	0.0339	0.6266	0.0518	0.0392	0.6479	0.0559
203907	0.0221	0.6519	0.0335	0.0253	0.6454	0.0348

Table 7.2
Convergence and upper bounds with $\Theta = 0.4$ and $\delta = 0.8367$.

DoFs	$\|A^{-1/2}(\sigma - \sigma_h)\|_0$	rate	$\mathcal{UB}(\sigma)$	$\|A^{1/2}\nabla(u - u_h)\|_0$	rate	$\mathcal{UB}(u)$
2092	0.2171	-	0.3383	0.2338	-	0.4228
2941	0.1686	0.7766	0.2523	0.1915	0.8191	0.2894
4495	0.1371	0.8132	0.2067	0.1570	0.8198	0.2267
6539	0.1172	0.8549	0.1775	0.1334	0.8497	0.1930
9962	0.0971	0.8285	0.1472	0.1107	0.8298	0.1567
14766	0.0787	0.8105	0.1204	0.0914	0.8257	0.1264

accuracy of upper bounds obtained in (4.3) for $\|A^{-1/2}(\sigma - \sigma_h)\|_0$ and (4.4) for $\|A^{1/2}\nabla(u - u_h)\|_0$. Let $\Omega = [-1, 1] \times [-1, 1]$ and we consider the following model problem

$$-\nabla \cdot A\nabla u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega,$$

where $A = \begin{bmatrix} 2 + \sin(xy) & 0 \\ 0 & 1 \end{bmatrix}$ with the true solution $u = (x^2 + y^2)^{0.51}(1 - x^2)(1 - y^2)$.

For the approximation spaces, we use the lowest-order Raviart-Thomas spaces for the flux variable $\sigma = -A\nabla u$ and the standard continuous piecewise linear functions for the primary variable u. With the reduction rate for elements $\gamma = \frac{1}{2}$, we choose $\Theta = 0.8$ and $\Theta = 0.4$ for the bulk parameters. Our algorithm selects a subset $\mathcal{M}$ of $\mathcal{T}_h$ that satisfies $\Theta\zeta^2(\Omega) \le \zeta^2(\mathcal{M})$ and $\Theta\,\mathrm{osc}(f, \mathcal{T}_h) \le \mathrm{osc}(f, \mathcal{M})$. Then, the MATLAB function "refinement.m" is used to refine the current mesh by dividing each marked triangle into four triangles of the same shape. The reduction rate of the error for $\|A^{-1/2}(\sigma - \sigma_h)\|_0$ and $\|A^{1/2}\nabla(u - u_h)\|_0$ is close to $\delta = \sqrt{1 - \Theta(1 - \gamma^2)}$, and this is better than the rate of convergence expected from (6.20). Also, the guaranteed upper bounds are close to the actual errors, and they are overestimated by a factor less than 2.

Tables 7.1 and 7.2 present the convergence behavior of error $\|A^{-1/2}(\sigma - \sigma_h)\|_0$ and $\|A^{1/2}\nabla(u - u_h)\|_0$ and their upper bounds. Note that $\mathcal{UB}(\sigma)$ and $\mathcal{UB}(u)$ are the upper bounds of $\|A^{-1/2}(\sigma - \sigma_h)\|_0$ and $\|A^{1/2}\nabla(u - u_h)\|_0$ defined in (4.3) and (4.4) respectively. Fig. 7.1 shows the surface plot of the true solution, and Fig. 7.2 shows the initial and adaptive meshes. Moreover, we present the contour plot of the approximation solution u_h in Fig. 7.3.

Data availability

No data was used for the research described in the article.

Acknowledgements

The author would like to thank two anonymous referees for their careful reading of this paper and their many helpful suggestions for improving the presentation of the results.

References

[1] M. Ainsworth, J.T. Oden, A Posteriori Error Estimation in Finite Element Analysis, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000.
[2] M. Bebendorf, A note on the Poincaré inequality for convex domains, Z. Anal. Anwend. 22 (2003) 751–756.
[3] P. Binev, W. Dahmen, R. DeVore, Adaptive finite element methods with convergence rates, Numer. Math. 97 (2004) 219–268.
[4] P.B. Bochev, M.D. Gunzburger, Least-Squares Finite Element Methods, Applied Mathematical Sciences, vol. 166, Springer, New York, 2009.
[5] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.
[6] F. Brezzi, J. Douglas Jr., L.D. Marini, Two families of mixed finite elements for second order elliptic problems, Numer. Math. 47 (1985) 217–235.

[7] P. Bringmann, C. Carstensen, G. Starke, An adaptive least-squares FEM for linear elasticity with optimal convergence rates, SIAM J. Numer. Anal. 56 (2018) 428–447.
[8] C. Carstensen, Collective marking for adaptive least-squares finite element methods with optimal rates, Math. Comput. 89 (2020) 89–103.
[9] C. Carstensen, R.H.W. Hoppe, Error reduction and convergence for an adaptive mixed finite element method, Math. Comput. 75 (2006) 1033–1042.
[10] C. Carstensen, E.-J. Park, Convergence and optimality of adaptive least squares finite element methods, SIAM J. Numer. Anal. 53 (2015) 43–62.
[11] C. Carstensen, E.-J. Park, P. Bringmann, Convergence of natural adaptive least squares finite element methods, Numer. Math. 136 (2017) 1097–1115.
[12] C. Carstensen, J. Storn, Asymptotic exactness of the least-squares finite element residual, SIAM J. Numer. Anal. 56 (2018) 2008–2028.
[13] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, SIAM J. Numer. Anal. 33 (1996) 1106–1124.
[14] T. Führer, D. Praetorius, A short note on plain convergence of adaptive least-squares finite element methods, Comput. Math. Appl. 80 (2020) 1619–1632.
[15] D. Gallistl, E. Süli, Mixed finite element approximation of the Hamilton-Jacobi-Bellman equation with Cordes coefficients, SIAM J. Numer. Anal. 57 (2019) 592–614.
[16] J. Ku, A posteriori error estimates for the primary and dual variables for the div first-order least-square finite element method, Comput. Methods Appl. Mech. Eng. 200 (2011) 830–836.
[17] P. Morin, R.H. Nochetto, K.G. Siebert, Data oscillation and convergence of adaptive FEM, SIAM J. Numer. Anal. 38 (2000) 466–488.
[18] P. Morin, R.H. Nochetto, K.G. Siebert, Convergence of adaptive finite element methods, SIAM Rev. 44 (2002) 631–658, Revised reprint of "Data oscillation and convergence of adaptive FEM", SIAM J. Numer. Anal. 38 (2) (2000) 466–488 (electronic), MR1770058 (2001g:65157).
[19] P. Morin, K.G. Siebert, A. Veeser, A basic convergence result for conforming adaptive finite elements, Math. Models Methods Appl. Sci. 18 (2008) 707–737.
[20] S. Nicaise, E. Creusé, Isotropic and anisotropic a posteriori error estimation of the mixed finite element method for second order operators in divergence form, Electron. Trans. Numer. Anal. 23 (2006) 38–62.
[21] P.-A. Raviart, J.M. Thomas, A mixed finite element method for 2nd order elliptic problems, in: Mathematical Aspects of Finite Element Methods, Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975, in: Lecture Notes in Math., vol. 606, Springer, Berlin, 1977, pp. 292–315.
[22] L.R. Scott, S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions, Math. Comput. 54 (1990) 483–493.
[23] R. Verfürth, Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation, Numer. Math. 78 (1998) 479–493.